RelationVis: Visual Analytics of Multifaceted Relationships between Different Entities in Text Collections

Shaocong Tan[1] Sihang Li[1] Shuai Chen[1] Jiacheng Yu[1] Chuanming Huang[3] Liang Tang[3]
Qinghua Shang[3] Zixi Fu[3] Zhuo Zhang[3] Xiaoru Yuan[1, 2]

1) Key Laboratory of Machine Perception (Ministry of Education), and School of EECS, Peking University
2) National Engineering Laboratory for Big Data Analysis Technology and Application, Peking University
3) QI-ANXIN Technology Group Inc.

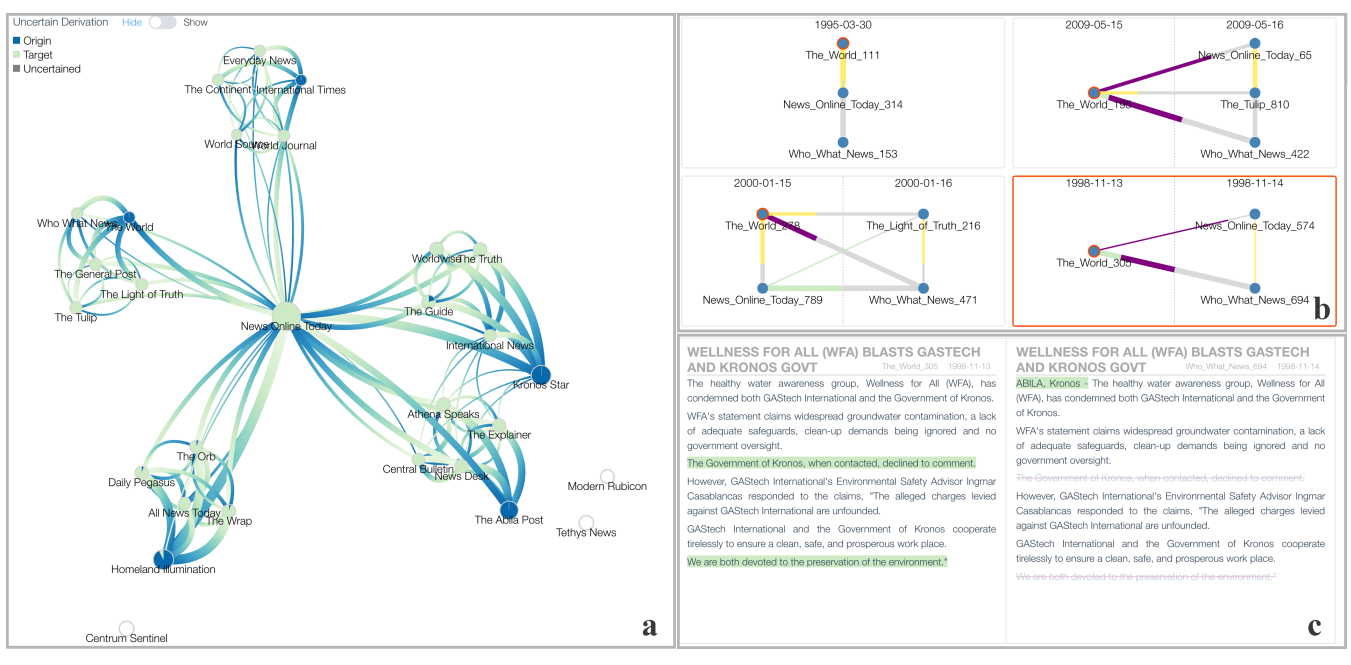
Introduction

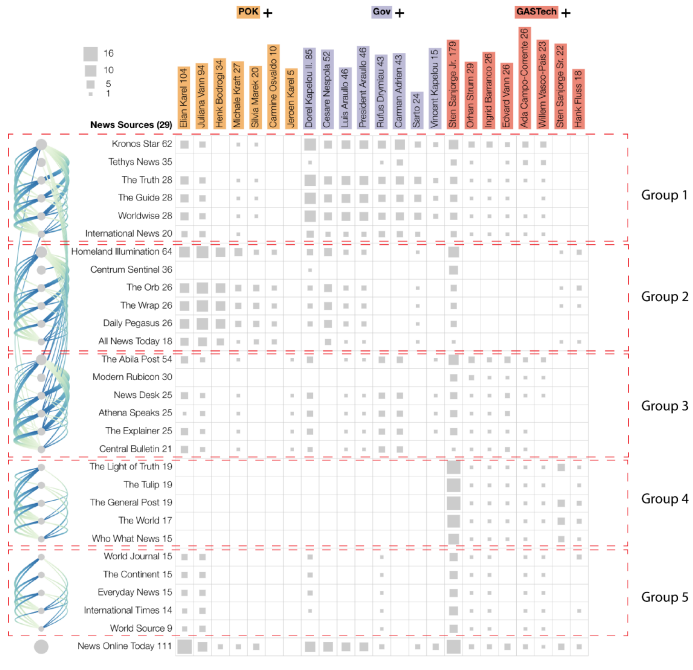
We propose RelationVis, a visual analytics system for exploring relationships between entities. We design the dissemination view to identify the primary and derivative sources, the matrix-based bias view to reveal their focal points, and the entity graph to display relationships between organizations and individuals mentioned in the collections. Our system provides overview and cross-linking interactions, supporting the user to mine relations from different perspectives.

Background

Text collections contain a high variety of relationships, whose keynote and phraseology represent the position on entities. In Mini-Challenge 1, there is a collection of 845 news, published by 29 media from 1982 to 2014, whose sources, authors, titles, contents and publication dates are available. We also get resumes of 35 employees and headers of 1175 internal mails.

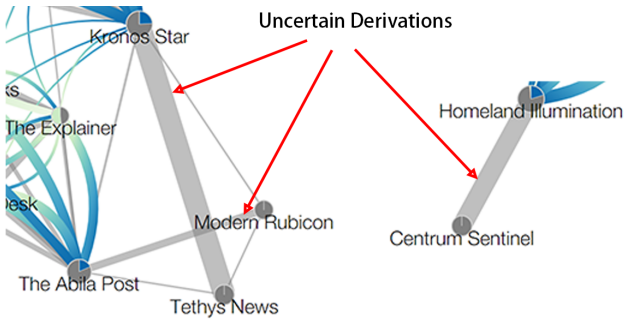
System Overview





Dissemination view

(a) We vectorize news by TF-IDF and generate clusters based on cosine distances. The latent derivation relationships between news sources are visualized by directed edge of the node-link diagram. The pie chart embedded in the node shows the percentage of original news.



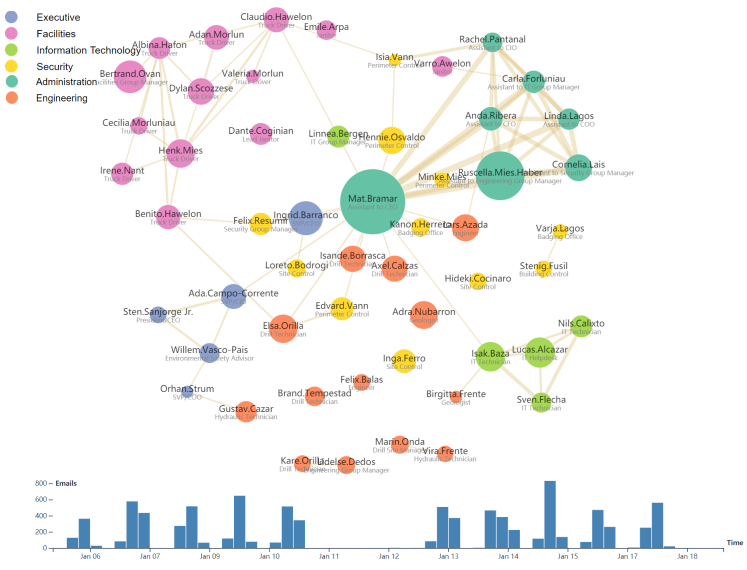
(b) The small multiple shows the abstract of dissemination, where each card is a cluster. The ratio of purple, green and yellow on edge encodes the proportion of deletion, addition and modification, while the thickness of edge encodes the text similarity between articles.
(c) After clicking an edge, the particulars of article modification will be shown in the text alignment view with colored words.

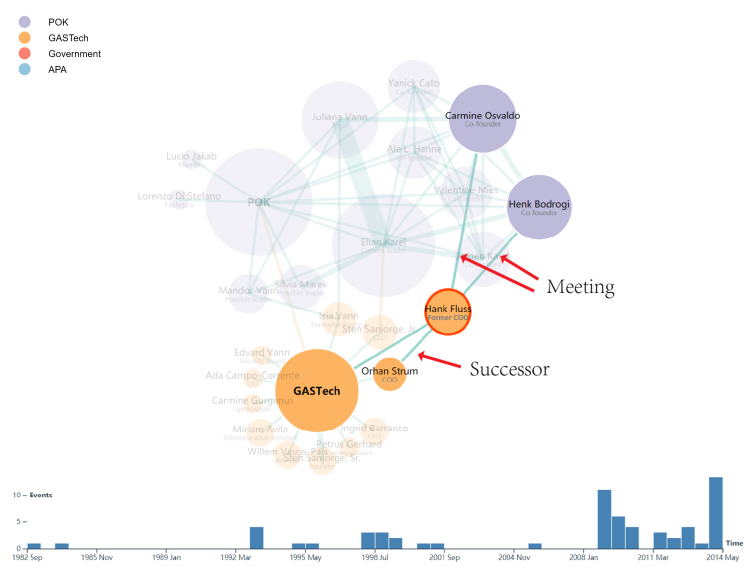
Bias view

The matrix-based view shows the coverage, gatekeeping, and statement biases of news sources against different entities. The square in the cell encodes the number of news reported by the source that covers the entity. The matrix can be expanded as parallel tag clouds to show topics. Thereafter, users can select a keyword to investigate how the news talks about the entity.

Entity view

The node-link diagrams represent the relationships between entities directly, including organization and individual. The thickness of the edge encodes the strength of the relationships, i.e., email contacts or occurrence in news. The color labels the property of relations calculated by the subject of news, green for the allies and yellow for the opponents. The timeline serves as a filter for selecting a period of interest to explore the evolution of relations.





Acknowledgement

This work is funded by the National Natural Science Foundation of China (No. 61872013).

Contact

xiaoru.yuan@pku.edu.cn
http://vis.pku.edu.cn

VisPKU